INSTITUTO SUPERIOR DE ECONOMIA E GESTÃO

UNIVERSIDADE TÉCNICA DE LISBOA

COMPARAÇÃO DO COMPORTAMENTO DE TESTES DE ESPECIFICAÇÃO EM MODELOS LOGIT MULTINOMIAIS

Maria da Conceição Torres Figueiredo

Lisboa, Julho de 1994

INSTITUTO SUPERIOR DE ECONOMIA E GESTÃO

UNIVERSIDADE TÉCNICA DE LISBOA

COMPARAÇÃO DO COMPORTAMENTO DE TESTES DE ESPECIFICAÇÃO EM MODELOS LOGIT MULTINOMIAIS

Dissertação apresentada como requisito parcial para obtenção do grau de mestre em Matemática Aplicada à Economia e à Gestão

Maria da Conceição Torres Figueiredo

Lisboa, Julho de 1994

Ao meu filho,

Tiago

AGRADECIMENTOS

Ao orientador desta dissertação, professor doutor João Caravana Santos Silva, expresso o meu mais sincero agradecimento por todo o seu apoio. Há que realçar a sua compreensão, a sua disponibilidade e os seus conhecimentos que, ao longo da realização deste trabalho, me foram indispensáveis.

À JNICT, Junta Nacional de Investigação Científica e Tecnológica, agradeço os meios financeiros colocados ao meu dispôr, através de uma bolsa de estudo, que permitiu a possibilidade de obter o grau de mestre.

Agradeço também, de um modo geral, a todos aqueles que sempre se mostraram prestáveis para ajudar à concretização deste meu objectivo.

INDÍCE

1. INTRODUÇÃO

Um dos desenvolvimentos mais importantes nas últimas décadas têm sido os modelos de resposta qualitativa, cuja característica principal é a variável aleatória endógena (ou dependente) assumir apenas valores discretos. Daí também a designação de modelos de escolha discreta.

Dentro destes, o mais utilizado é o modelo logit dada a facilidade em ser extensível a mais do que duas alternativas e devido à versatilidade da sua especificação relativamente às variáveis explicativas e aos seus efeitos. Por outro lado é particularmente atractivo para os economistas, dada a sua interpretação como um modelo de escolha discreta devido à maximização da utilidade.

Contudo é um modelo pouco testado, e atendendo ao facto de que é bastante sensível dadas as hipóteses em que assenta, esta tese pretende ser um estudo da comparação do comportamento de testes de especificação em modelos logit multinomiais (MLM), procurando de alguma forma explorar e dar continuidade ao estudo levado a cabo por Fry e Harris (1993): "A Monte Carlo Study of Tests for the Independence of Irrelevant Alternatives Property".

Assim, no ponto dois da dissertação, é descrito o modelo de escolha discreta, os pressupostos em que assenta.

No ponto três, apresenta-se o modelo logit multinomial, definindo e questionando a propriedade a ser testada, ou seja, a independência das alternativas irrelevantes (IAI).

Seguidamente, no ponto quatro são desenvolvidos alguns testes de especificação, que se dividem em dois grandes grupos:

• um, baseado na partição do conjunto de escolha - fazendo apelo ao teste de Hausman (1978) - onde se a propriedade que é testada se verificar, então a estrutura do modelo e os parâmetros não se alteram quando a escolha incide sobre um subconjunto do conjunto considerado;

• o outro, baseado na especificação de um modelo alternativo, que não goze da propriedade testada, em que se considera uma generalização do modelo logit multinomial. Como este é um caso particularde um modelo mais geral, os testes clássicos (teste do multiplicador de Lagrange, da razão de verosimilhança e do Wald) podem ser aplicados.

O ponto cinco, constituí a segunda parte desta dissertação onde se faz um estudo de simulação sobre três testes, o que apresenta melhores resultados no artigo referido e outros dois não considerados, nomeadamente o teste do multiplicador de Lagrange contra o modelo logit sequencial e o teste da matriz de informação contra o modelo logit heterogéneo, de modo a avaliá-los e compará-los em termos de potência e dimensão.

No ponto seis, serão tiradas algumas conclusões, tendo sempre presente o artigo em causa, que permitirão optar, sob determinadas condições, por um dos testes estudados.

2. MODELOS DE ESCOLHA DISCRETA

Os modelos de resposta discreta ou qualitativa são modelos de regressão caracterizados por uma variável dummy dependente que indica qual a alternativa escolhida de entre um conjunto finito de alternativas. O que leva a que estes modelos tenham bastante aplicação na economia, dada a natureza das respostas relativas ao comportamento dos indivíduos. Ou seja, a economia, como teoria de escolha, não é só aplicada a questões de quanto produzir ou consumir mas também se se produz ou consome determinado bem. São situações em que o decisor tem de escolher uma "acção" dentro de um conjunto finito de alternativas discretas. Na maior parte das vezes é uma decisão do tipo sim ou não: por exemplo, aceita ou não aceita determinado emprego, compra ou não compra carro, nestes casos estamos perante modelos binomiais [ver exemplos no survey de Amemiya (1981)].

Para estudar este tipo de modelos assume-se que os decisores têm um comportamento racional, baseando-se nas opiniões, convicções acerca da decisão que tomam. Apesar do conceito de racionalidade ser bastante ambíguo, pode estabelecer-se um determinado conjunto de regras de modo a atenuar tal problema: assim, define-se como decisor racional o que toma decisões com base em preferências transitivas (se a alternativa 1 é preferível à alternativa 2 e esta preferível à alternativa 3, então a alternativa 1 é preferível à 3) e consistentes (i.e., sob condições idênticas o decisor tem sempre a mesma preferência).

Considera-se um processo de decisão seguro e consistente, na medida que os indivíduos seguem os seus próprios objectivos, quaisquer que sejam, sendo postas de lado as situações em que reagem de modo diferente, de acordo com o seu estado de espírito na altura (impulsividade).

Os indivíduos são colocados perante um conjunto finito de alternativas, designado por C, onde as características de cada alternativa i podem ser comensuráveis, ou seja, a atractividade de uma alternativa expressa por um vector de valores das características é reduzível a um escalar. Então pode definir-se uma única função objectivo que expressa a utilidade de uma alternativa, em termos das suas características. Por hipótese, essa utilidade está relacionada com uma escala de preferências ordinal, i.e., apenas se pode ordenar as preferências por cada alternativa, através de relações de maior que, menor que ou de igual. Contudo, quando necessário, é substituido pelo conceito de utilidade cardinal [ver Ben-Akiva e Lerman (1985)].

Enquanto que o indivíduo é suposto escolher sempre a alternativa com maior utilidade, o analista devido a factores que Manski (1973) identificou como características e variações de gosto não observadas, erros de medida e uso de variáveis instrumentais (ou proxi), tem de tratar a utilidade como uma variável aleatória.

Definindo então, a utilidade da alternativa i como:

$$U_i = V_i + \varepsilon_i \quad , \forall i \in C$$

sendo V_i a componente sistemática e ε_i a componente aleatória.[1]

Para cada indivíduo, qualquer alternativa i pode ser caracterizada por um vector de atributos (Z_i) e por características do próprio indivíduo (S).

---

[1] v_i é uma função determinística (i.e. não aleatória) , ao passo que o termo ε_i pode também ser uma função mas é aleatório do ponto de vista do analista.

Logo $X=[Z_i,S]\Rightarrow V_i=V(X_i)$

Sendo β o vector dos parâmetros desconhecidos, a utilidade da alternativa i pode ser expressa como:

$$U_i=V(X_i,\beta)+\varepsilon_i$$

Tendo em conta que o indivíduo racional vai maximizar a sua função utilidade [ver Ben-Akiva e Lerman (1985)], pode-se definir a probabilidade de qualquer alternativa i ser escolhida, como a probabilidade da utilidade da alternativa i ser maior ou igual às utilidades de todas as outras alternativas pertencentes ao conjunto C :

$$P(i|C)=\Pr(U_i\geq U_j) \quad ,\forall j\neq i \tag{1}$$

Ou seja,

$$P(i|C)=\Pr(V_i+\varepsilon_i\geq V_j+\varepsilon_j) \quad ,\forall j\neq i \tag{2}$$

Definida a componente sistemática V_i, falta a componente aleatória. Dependendo das hipóteses que se façam acerca da distribuição dos erros, assim se obtêm diferentes modelos de escolha discreta.

Não se pense, todavia, que a especificação da distribuição dos ε_i's é independente da especificação dos V_i's, na medida que estes, ao refletirem todo o tipo de erros de observação discutidos atrás, levam a que diferentes especificações de V conduzirão a diferentes distribuições dos ε's .

Dentro destes modelos de escolha discreta, os mais complexos são aqueles em que a variável dependente assume mais do que dois valores discretos, designados por modelos multinomiais, apresentados pela primeira vez por Theil (1969) e desenvolvidos em seguida por McFadden (1974). Está-se perante o caso em que o decisor é confrontado com um conjunto J$\geq$ 2 de alternativas face às quais terá que escolher uma, baseado na utilidade esperada de cada alternativa.

Em termos econométricos significa que, não basta especificar a distribuição univariada das diferenças dos erros$(\varepsilon_j - \varepsilon_i)$, caso dos modelos binomiais, é igualmente necessário caracterizar a distribuição conjunta de todos os erros $f(\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_J)$. Para especificar um modelo de utilidade aleatória, supõe-se inicialmente que se dispõe de uma probabilidade conjunta para os erros.

3. MODELO LOGIT MULTINOMIAL

3.1 Especificação e Hipóteses do Modelo

A hipótese base deste modelo diz respeito à especificação da distribuição dos erros, tal como ficou anteriormente estabelecido. Se bem que seja desejável que a distribuição dos erros seja o mais flexível possível, há que ter em conta que tal vai dificultar a estimação.

No caso do modelo logit multinomial, ao impôrem-se hipóteses bastante restritivas quanto aos erros, sacrifica-se a flexibilidade em troca da facilidade de estimação.

Assim, assume-se que as variáveis aleatórias ε_i's são independentes e identicamente distribuidas, seguindo uma distribuição Gumbel (Valor Extremo Tipo I) [ver Johnson e Kotz (1970)].

Isto é,

$$F(\varepsilon_i) = \exp\{-\exp[-\mu_i(\varepsilon_i - \eta_i)]\}$$

em que η_i é um parâmetro de localização e μ_i um parâmetro de escala positivo. Considerando-se por um lado, μ_i igual para todas as alternativas e indivíduos (Hipótese de Homocedasticidade da Componente Aleatória) pode-se, sem perda de generalidade, fazer $\mu_i = 1$, o que implica variâncias iguais da componente aleatória[2]; e por outro lado, η_i constante, considerando-se, por exemplo $\eta_i = 0$, o que não é nenhuma hipótese restritiva, na medida que a utilidade sistemática tenha um termo constante.

Estas hipóteses levam a que

$$F(\varepsilon_i) = \exp[-\exp(-\varepsilon_i)]$$

Então, demonstra-se [ver Domencich e McFadden (1975) e Ben-Akiva e Lerman (1985)] que a expressão (2) que representa a probabilidade da alternativa i (condicionada às variavéis explicativas X_i) ser escolhida, pode ser escrita como:

$$P(i|C) = \frac{\exp V_i}{\sum_j \exp V_j} \tag{3}$$

Como por hipótese V_i é combinação linear dos elementos de X_i (Hipótese de Linearidade) a expressão anterior vem:

$$P(i|C) = \frac{\exp(X_i\beta)}{\sum_j \exp(X_j\beta)} \tag{4}$$

Uma das hipóteses mais fortes e polémicas relativamente a este tipo de modelos é os parâmetros do vector serem constantes em relação às alternativas e aos indivíduos. Se quanto às alternativas não é rigído, podendo haver diferentes coeficientes para diferentes alternativas através da inclusão de varíaveis dummy, já em relação aos indivíduos considerar os β's constantes significa que toda a população tem os mesmos

---

[2]A variância da distribuição Gumbel é igual a $\frac{\pi^2}{6\mu_i^2}$ ficando $\frac{\pi^2}{6}$.

gostos, o que pode ser questionável em determinadas circunstâncias, podendo o modelo logit multinomial captar variações de gosto que são funções de variáveis observadas, levantando o problema de os parâmetros do modelo variarem só com alguma variável não observada ou aleatoriamente.

Por questões de identificação é necessário normalizar os parâmetros, por exemplo, considerar $X_1 = 0$, o que origina que a expressão (4) se transforme em

$$P(1) = \frac{1}{1 + \sum_{j=2}^{J} \exp(X_j \beta)} \tag{5}$$

$$P(i) = \frac{\exp(X_i \beta)}{1 + \sum_{j=2}^{J} \exp(X_j \beta)} \qquad i = 2, \ldots, J \tag{6}$$

A distribuição dos erros do modelo logit multinomial conduz-nos a dois factos importantes:

- A facilidade em estimar-se o vector dos parâmetros β através do método da máxima verosimilhança;

- A imposição de uma propriedade bastante restritiva e controversa, a Independência das Alternativas Irrelevantes (IAI).

3.2. Estimação do Modelo

Partindo da hipótese de que os dados da amostra são obtidos através da realização de provas aleatórias independentes, as quais podem ser com ou sem reposição, a função de verosimilhança de um modelo logit multinomial pode ser expressa como:

$$L = \prod_{i \in C} P(i)^{Y_i} \tag{7}$$

Sendo Y_i uma variável binária tal que:

$$Y_i = \begin{cases} 1 & \text{se a alternativa i for escolhida} \\ 0 & \text{caso contrário} \end{cases}$$

Substituindo as expressões (5) e (6) na função de verosimilhança (7) obtém-se

$$L = \frac{1}{1 + \sum_{j=2}^{J} \exp(X_j \beta)} \prod_{j=2}^{J} \exp(X_j \beta)^{Y_i} \tag{8}$$

Tomando o logaritmo da função de verosimilhança, vem

$$\ln L = \sum_{j=2}^{J} Y_i X_j \beta - \ln\left[1 + \sum_{j=2}^{J} \exp(X_j \beta)\right] \tag{9}$$

Como a função de verosimilhança é não linear nos parâmetros deve aplicar-se o método de Newton-Raphson que exige o cálculo das derivadas de primeira e segunda ordem do logaritmo da função de verosimilhança relativamente aos β's [ver Fomby (1984)].

McFadden (1974) mostra que, sob determinadas condições relativamente fracas, a função L na equação (8) é globalmente côncava (o que facilita a procura do máximo relativamente aos parâmetros), logo a solução existe e é única. Então o estimador da máxima verosimilhança do vector é consistente, assimptoticamente normal e assimptoticamente eficiente.

A interpretação dos coeficientes estimados é diferente da que é feita no caso dos modelos de regressão linear, ou seja, não indicam a variação da probabilidade de ocorrência de determinado acontecimento dada uma variação unitária na variável independente correspondente, mas refletem os efeitos das variações da variável independente no $\log[P(i)/1-P(i)]$. A "grandeza" desses efeitos depende da probabilidade original e assim dos valores iniciais de todas as variáveis independentes e dos seus coeficientes.

3.3. Elasticidades

A elasticidade define a sensibilidade da probabilidade de escolha de um indivíduo a uma variação do valor de qualquer característica.

O caso mais simples é a elasticidade da probabilidade de um indivíduo escolher a alternativa i relativamente a uma variação de uma qualquer característica (variável independente do modelo) nomeadamente uma das x_{ik}'s.

Assim a elasticidade directa do logit é dada por:

$$\mathrm{E}_{x_{jk}}^{P(i)} = \frac{\partial P(i)}{\partial x_{jk}} \cdot \frac{x_{ik}}{P(i)} = \frac{\partial \ln P(i)}{\partial \ln x_{ik}} = [1 - P(i)] x_{ik} \beta_k$$

Analogamente, a elasticidade cruzada da probabilidade da alternativa i ser escolhida relativamente a uma característica da alternativa j, define-se como:

$$E_{x_{jk}}^{P(i)} = \frac{\partial \ln P(i)}{\partial \ln x_{jk}} = -P(j)x_{jk}\beta_k \quad para \quad j \neq i$$

Uma das propriedades do modelo logit multinomial é ter elasticidades cruzadas uniformes, ou seja, as elasticidades cruzadas de todas as alternativas em relação à variação de uma característica (que afecta só a utilidade da alternativa j) são iguais para todas as alternativas $j \neq i$.

No fundo, esta propriedade não passa de uma "manifestação" de uma das propriedades mais controversas destes modelos: a Independência das Alternativas Irrelevantes (IAI).

3.4. Prop. da Independência das Alternativas Irrelevantes (IAI)

3.4.1. Definição

Esta propriedade resulta da hipótese referida como o "axioma de escolha", que é a base para a construção do modelo de utilidade constante desenvolvido Luce (1959)[3]:

Um conjunto de alternativas, definido como qualquer subconjunto de C, satisfaz o "axioma de escolha" contando que para todo o i, $\tilde{C}$ e C, tal que $i \in \tilde{C} \subseteq C$, $P(i|\tilde{C} \subseteq C) = P(i|\tilde{C})$, desde que a probabilidade condicional exista[4].

---

[3]Para mais detalhes ver Ben-Akiva (1983)

[4]Desde que $0 \leq P(i|\tilde{C}) \leq 1 \ \forall i \in \tilde{C}$ e $\sum_{i \in \tilde{C}} P(i|\tilde{C}) = 1$

Por outras palavras, se algumas alternativas forem retiradas ao conjunto de escolha, a probabilidade de escolha relativa do conjunto C não se altera. As probabilidades de escolha de um subconjunto de alternativas só dependem de si próprias, sendo independentes de outras quaisquer alternativas que possam existir .

Substituindo a expressão do axioma na definição de probabilidade condicional de escolha[5] obtém-se

$$P(i|C) = P(i|\tilde{C})P(\tilde{C}|C) \qquad i \in \tilde{C} \subseteq C$$

Ou seja, a probabilidade de escolher a alternativa i de C é igual ao produto da probabilidade de escolher a alternativa i pertencente ao subconjunto $\tilde{C}$ pela probabilidade da escolha recair numa alternativa do mesmo subconjunto.

Pode-se assim, obter a fórmula da propriedade IAI,

$$\frac{P(i|\tilde{C})}{P(j|\tilde{C})} = \frac{P(i|C)}{P(j|C)}, \qquad i, j \in \tilde{C} \subseteq C$$

O que significa que para um determinado indivíduo, a razão das probabilidades de escolha de duas alternativas não é afectada pelas utilidades sistemáticas de outras quaisquer.

[5] $P(i|\tilde{C} \in C) = P(i|C)/P(\tilde{C}|C)$ onde $P(\tilde{C}|C) = \sum_{i \in C} P(i|C) = 1$

Logo, como se pode ver:

$$\frac{P(i)}{P(l)} = \frac{\exp V_i \Big/ \sum_j \exp V_j}{\exp V_l \Big/ \sum_j \exp V_j} = \frac{\exp V_i}{\exp V_l} = \frac{\exp X_i \beta}{\exp X_l \beta} = \exp(X_i - X_l)\beta$$

Se é uma propriedade desejável nos modelos determinísticos de escolha de consumo (i.e., em termos do próprio indivíduo), as dúvidas levantam-se ao tratar de modelos de escolha probabilística (em termos do analista), ou seja, os problemas aparecem quando as alternativas e os indivíduos não estão completamente caracterizados e portanto há que ter em conta a probabilidade de um indivíduo seleccionado aleatoriamente escolher uma qualquer alternativa.

3.4.2. Implicações

Apesar de ser considerada a principal limitação do modelo logit multinomial, é extremamente útil em termos de previsão, na medida que, sob determinadas condições, permite prevêr escolhas face a alternativas não disponíveis aquando a estimação, já que, a probabilidade de escolher a alternativa i (sendo certo que uma das alternativas, i ou j, é escolhida), condicionada às características de todas as alternativas disponíveis, depende somente das características destas duas alternativas.

A validade desta propriedade depende da estrutura do conjunto de escolha, isto é, só se verifica em conjuntos de alternativas distintos (com erros não correlacionados) e ponderados independentemente do ponto de vista do decisor, tal como é ilustrado pelo Paradoxo do autocarro azul /autocarro vermelho [ver Ben-Akiva e Lerman (1985)], em que existem alternativas que são substitutos perfeitos.

Tal limitação fica a dever-se à hipótese de os erros serem mutuamente independentes afastando qualquer possibilidade de o modelo captar a situação em que duas alternativas, similares ou não, terem erros semelhantes.[6]

Outra característica, é o facto da IAI só se aplicar a grupos homogéneos de indivíduos e não à população como um todo, ou seja, as probabilidades dos indivíduos não se transpõem para as quotas da população, para tal há que específicar cuidadosamente a componente sistemática da função utilidade, de modo a reflectir a heterogeneidade característica de uma população.

Esta propriedade restringe bastante a flexibilidade da forma funcional, já que obriga a iguais elasticidades cruzadas das probabilidades de várias alternativas relativamente a uma característica de determinada alternativa.

Estas limitações, quando não tidas em conta, podem levar a previsões erradas. Daí, esta propriedade ser considerada por vezes como inapropriada, apesar de se manter como hipótese subjacente em muitas aplicações.

Outro aspecto essencial que convém sublinhar, é esta propriedade só se verificar se V_i depender somente de X_i , i.e., se for independente do conjunto de escolha (o vector x_i para a alternativa i é independente das características das outras alternativas). Logo, esta propriedade não é "intrínseca" a este tipo de modelos, mas sim consequência de uma restrição. Onde, caso não se verifique, qualquer modelo pode ser expresso na forma de um logit, já que a sua forma é suficientemente flexível para aproximar qualquer modelo

---

[6] Não são tratados os diferentes graus de substituibilidade ou complementariedade entre as alternativas.

de probabilidade , daí se falar do modelo logit universal [ver Cramer e Ridder (1988)].

Pode ser visto, notando que:

$$P(i) = f(X_i\ldots) \qquad P(i) = \frac{f(X_i\ldots)}{\Sigma f(X_j\ldots)} = \frac{\exp[\log f(X_i\ldots)]}{\Sigma \exp[\log f(X_j\ldots)]}$$

$$\sum_{j=1}^{J} f(X_j\ldots) = 1 \qquad f(X_j\ldots) > 0$$

Então a hipótese do modelo logit multinomial referente à distribuição dos erros, não é tão forte como pode parecer à primeira vista. Ou seja, a forma de probabilidade do logit multinomial não é particularmente restritiva.

3.5. Modelo Probit Multinomial

O modelo probit multinomial é menos restritivo que o modelo logit multinomial. Onde a diferença fundamental reside na hipótese da distribuição dos resíduos. Enquanto no modelo logit multinomial estes seguem uma distribuição Gumbel, no modelo probit multinomial seguem uma distribuição Normal Multivariada com um vector de médias 0 e uma matriz de covariâncias $\Sigma_\varepsilon(J \times J)$.

O modelo probit multinomial (MPM) supera assim, a "fraqueza" do logit multinomial, ao permitir que os erros sejam correlacionados (o que significa que não têm de ser independentes) e que os parâmetros β variem normalmente em relação aos indivíduos.

O modelo de utilidade aleatória é expresso por:

$$U_{it} = X_{it}\beta_{it} + \varepsilon_{it}$$

sendo $\beta_{kt} = \bar{\beta} + v_{kt}$, onde $\bar{\beta}$ é um parâmetro de gosto médio e v_{kt} é um vector de elementos aleatórios que representam os desvios do t-ésimo indivíduo em relação à média.

Sendo assim, a independência das componentes estocásticas da utilidade não implica a propriedade IAI, já que tem em conta a similaridade das alternativas.

O grande inconveniente é a sua computação envolver o cálculo de integrais mũltiplos (por exemplo para um modelo de J alternativas a dimensão do integral é J-2 alternativas), assim, só pode ser aplicado para um número reduzido de alternativas (de preferência até quatro).

Para ilustrar tal complexidade, considere-se o caso de J=3. Assim, para calcular, por exemplo, P(Y=2) temos de calcular o integral múltiplo:

$$P(Y=2) = P(U_2 \geq U_1, U_2 \geq U_0)$$

$$= \int_{-\infty}^{\infty} \int_{-\infty}^{U_2} \int_{-\infty}^{U_2} f(U_0, U_1, U_2) dU_0 dU_1 dU_2$$

em que $f(U_0, U_1, U_2)$ é a função densidade de probabilidade da distribuição normal trivariada.

Se bem que, foram sugeridos dois métodos para aproximar o MPM a probabilidades de escolha com um custo razoável: o Método Monte Carlo usado por Lerman e Manski (1982) e outro baseado na aproximação de Clark (1961).

McFadden (1989) desenvolveu o método dos momentos simulados propondo uma modificação simples do método dos momentos, em que substitui as probabilidades $P(i)$ que necessitam de integração numérica por estimadores obtidos através de simulações de Monte Carlo. Este método baseia-se no facto de que, apesar de, a função de probabilidade ser difícil de expressar analitíca ou computacionalmente, é relativamente fácil de simular. Quando esta função é substituida por um simulador não enviesado de tal modo que os erros de simulação são independentes em relação às alternativas, a variância introduzida pela simulação será controlada através da Lei dos Grandes Números. Sendo desnecessário estimar consistentemente cada resposta esperada.

O modelo probit multinomial fornece assim uma especificação alternativa dos modelos de escolha discreta, sem qualquer necessidade da IAI, permitindo que os erros estejam correlacionados em relação às alternativas, que as variâncias dependam das alternativas e variações de gosto aleatórias relativamente aos indivíduos. Contudo é ainda um modelo restritivo, na medida que exige que a distribuição dos erros seja simétrica.

4. TESTES DE ESPECIFICAÇÃO

Na última década os testes de especificação têm registado um franco desenvolvimento. A ideia de que um modelo deve ser testado antes de ser considerado e utilizado como base adequada para um determinado estudo económico tem-se generalizado.

Os testes de especificação desempenham um importante papel na estimação de modelos econométricos, de tal modo que a sua necessidade tem sido realçada por muitos autores, nomeadamente Hendry (1980) sugere que as três regras de ouro dos econometristas são "testar,testar e testar".

Nos modelos de escolha discreta as hipóteses referentes à própria estrutura do modelo assentam nas propriedades da distribuição das utilidades aleatórias ou nas propriedades das probabilidades de escolha.

O modelo logit multinomial é um dos modelos mais aplicados na análise de escolha discreta, atendendo principalmente às vantagens obtidas na especificação do modelo, na estimação (particularmente vantagens computacionais) e na previsão.

Mas tal é conseguido à custa de uma restrição bastante forte, subjacente a este tipo de modelos, a propriedade IAI, apesar de, na maior parte dos casos, ser considerada como implausível e irrealista. O que a não verificar-se implica uma má especificação do modelo, conduzindo a estimadores de máxima verosimilhança inconsistentes. Outra imposição do modelo é assumir que não há variações aleatórias de gosto, ou seja, as diferenças de gosto entre os indíviduos são captadas por variáveis socioeconómicas incluídas na especificação do modelo.

Sabemos à priori que estas hipóteses básicas só podem ser consideradas aproximações razoáveis de relações mais complexas. Logo deve ser testada em todas as aplicações, ver até que ponto é violada, e quando tal acontecer tentar solucionar com o intuito de obter um modelo aceitável.

E assim testar a propriedade IAI é testar desvios à própria especificação do modelo logit[7].

A abordagem directa para detectar violações, é através da comparação dos resultados da estimação de um modelo generalizado, isto é, que não verifique a hipótese básica que se quer testar, nomeadamente a propriedade IAI, com um modelo logit.

Esta abordagem revela-se difícil, dado os custos computacionais a que obriga a estimação de um modelo com uma estrutura menos restritiva como é o caso do modelo probit multinomial, com uma matriz de covariâncias não restrita e com variações de gosto aleatórias de indivíduo para indivíduo, que é a forma mais geral disponível. Logo há que se concentrar em testes que não requeiram a estimação de modelos probit[8] .

Os testes para a propriedade IAI dividem-se em duas categorias:

- uma, baseada na partição do conjunto de escolha, onde a ideia inerente é que se a propriedade IAI é válida então a estrutura do modelo e os parâmetros não se alteram quando a escolha é analisada sobre um subconjunto do conjunto de escolha total. Ou seja ver até que ponto os parâmetros estimados ou os logaritmos da verosimilhança

---

[7] A propriedade IAI resulta integralmente de os erros serem independentes e identicamente distribuidos, com distribuição Gumbel: $G(\varepsilon_i)=\exp[-\exp(-\varepsilon_i)]$

[8] Se bem que Horowitz (1981a) construiu um teste tipo razão de verosimilhança de um modelo logit contra a alternativa de um probit, sem ser necessário estimar os parâmetros deste,este teste insere-se numa classe de testes denominada por testes do multiplicador de Lagrange [Silvey (1959)].

maximizados do conjunto de escolha completo ($C=\{1,...,J\}$) e dum subconjunto especifico $\tilde{C}$ ($\tilde{C} \subseteq C$) são significativamente diferentes uns dos outros.

- a outra, baseada na especificação de um modelo alternativo que não verifique a propriedade IAI, em que a ideia principal é considerar uma generalização do MLM, por exemplo o modelo logit sequencial (MLN), entre outros a analisar. Como o modelo logit multinomial é um caso particular do modelo mais geral, os testes indirectos da IAI (teste do multiplicador de Lagrange, da razão de verosimilhança e do Wald) podem ser aplicados.

4.1. O Teste de Hausman e McFadden (HM)

O teste envolve a comparação de modelos logit estimados a partir de subconjuntos de alternativas do conjunto de escolha total.

No modelo logit multinomial (MLM) onde as variáveis explicativas da alternativa i são função apenas das características da própria alternativa, a IAI verifica-se. O que implica que a estrutura do modelo e os parâmetros não se alteram quando a escolha é analisada, condicionada por um conjunto restrito do conjunto de escolha total .

O teste baseia-se na eliminação de uma ou mais alternativas do conjunto para ver se o comportamento de escolha subjacente ao conjunto de escolha restrito obedece à IAI.

Para tal estimam-se os parâmetros desconhecidos dos dois conjuntos de escolha. Se estes são aproximadamente os mesmos então não se rejeita a especificação do MLM, ou seja, a hipótese nula de que a IAI se verifica.

Hausman e McFadden (1984) desenvolveram um teste assimptótico para esta comparação, usando a abordagem dos testes de especificação introduzida por Hausman (1978):

Considerando o modelo logit (linear nos parâmetros) para o conjunto de escolha total:

$$P(i|C) = \frac{\exp(X_i\beta)}{\sum_{j \in C} \exp(X_j\beta)}$$

E o modelo logit (linear nos parâmetros) para o conjunto de escolha restrito:

$$P(i|\tilde{C}) = \frac{\exp(W_i\theta)}{\sum_{j \in \tilde{C}} \exp(W_j\theta)}$$

sendo β o vector dos parâmetros, $\beta = (\gamma', \theta')'$, em que γ' é um subvector de $[(K-k)\times 1]$ componentes não identificáveis e θ' o subvector das $(k \times 1)$ componentes identificáveis.[9]

E X a matriz das variáveis explicativas das características da alternativa j e/ou das características do decisor que afecta a escolha da mesma, $X = [Z', W']'$, em que Z' corresponde às variáveis das componentes não identificáveis e W' às das componentes identificáveis.

[9] Com a ausência de certas alternativas do C em $\tilde{C}$, geralmente só é possível identificar um subvector do β. Este subvector não inclui os parâmetros específicos a cada alternativa das alternativas não incluídas no conjunto restrito.

Através da maximização dos logaritmos das respectivas funções de verosimilhança:

$$Máx.\ \ln L_0(\beta) = \sum_{t=1}^{T} \sum_{j\in C} Y_i \ln P(i|C)$$

e

$$Máx.\ \ln L_1(\theta) = \sum_{t=1}^{T_1} \sum_{j\in \tilde{C}} Y_i \ln P(i|\tilde{C})$$

obtêm-se estimativas consistentes de β e θ, $\hat{\beta}_0 = (\hat{\gamma}_0, \hat{\theta}_0)$ e $\hat{\theta}_1$ respectivamente.

Sob as hipóteses de regularidade [ver Hausman e McFadden (1984)], quando a especificação do logit é correcta, a hipótese nula traduz-se em :

$$p\lim\ (\hat{\theta}_1 - \hat{\theta}_0) = 0$$

Então o teste HM requer a utilização de uma forma quadrática, envolvendo a diferença entre os subvectores dos parâmetros estimados para ambas as hipóteses e a diferença das respectivas matrizes de variâncias e covariâncias. Hausman e McFadden mostram que sob a hipótese nula,

$$HM = (\hat{\theta}_1 - \hat{\theta}_0)'\left[\hat{cov}(\hat{\theta}_1) - \hat{cov}(\hat{\theta}_0)\right]^{-1}(\hat{\theta}_1 - \hat{\theta}_0)$$

segue uma distribuição assimptótica χ^2, com um número de graus de liberdade igual à característica da matriz $\left[\hat{cov}(\hat{\theta}_1) - \hat{cov}(\hat{\theta}_0)\right]$, ou seja o número de elementos do subvector de coeficientes que é identificável a partir do conjunto de escolha restrito, i.e., a dimensão de $\hat{\theta}_0$.

Quando esta matriz é singular, a sua inversa é substituida pela inversa generalizada, o que levanta alguns problemas computacionais.

Convém ainda notar, que este teste é bastante sensível às alternativas retiradas do modelo que se pretende ensaiar [ver Fry e Harris (1993)].

4.2. O Teste de Small e Hsiao (SH)

Este teste pretende ser uma "solução" aos problemas levantados pelo teste de McFadden, Tye e Train (1977) , denominado por MTT, e de Horowitz (1981b).

Nomeadamente o teste MTT consiste na comparação das estimativas dos subvectores de parâmetros obtidos com os conjuntos de escolha total e restrito, através de um ensaio da razão de verosimilhança clássico:

$$MTT = -2\left[\ln L_1(\hat{\theta}_0) - \ln L_1(\hat{\theta}_1)\right]$$

Small e Hsiao chegaram à conclusão que o teste é assimptóticamente enviesado, ou seja, tende a não rejeitar a hipótese nula, devido a considerar $\hat{\theta}_0$ um subvector constante, o que é equivalente a assumir que $\hat{\theta}_0$ e $\hat{\theta}_1$ são não correlacionados, não sendo possível tal verificar-se já que ambas os estimativas são obtidas com base em informação comum.

Ao passo que o teste de Horowitz ao utilizar amostras estritamente independentes tende a ser assimptoticamente enviesado para rejeitar a hipótese nula.

Assim Small e Hsiao propõem uma versão modificada do teste razão de verosimilhança (um teste assimptoticamente centrado) baseado:

a) numa partição aleatória da amostra em duas subamostras (assimptoticamente com a mesma dimensão), respectivamente A e B de tamanhos N^A e N^B;

b) numa definição mais complexa das estimativas dos parâmetros que servem para efectuar o teste, a partir de valores estimados para os parâmetros de acordo com as subamostras.

Dividida aleatoriamente a amostra em duas subamostras de dimensões N^A e N^B, estimam-se os parâmetros através da maximização do logaritmo das respectivas funções de verosimilhança:

$$Máx.\ \ln L_0^A(\beta) = \sum_{j \in N^A} Y_i \ln P(i|C)$$

$$Máx.\ \ln L_0^B(\beta) = \sum_{j \in N^B} Y_i \ln P(i|C)$$

Conduzindo aos seguintes vectores:

$$\hat{\beta}_0^A = (\hat{\gamma}_0^{A'}, \hat{\theta}_0^{A'}) \quad \text{e} \quad \hat{\beta}_0^B = (\hat{\gamma}_0^{B'}, \hat{\theta}_0^{B'})$$

em que $\hat{\gamma}_0^{A'}$ e $\hat{\gamma}_0^{B'}$ correspondem aos subvectores de $[(K-k)\times 1]$ componentes não identificáveis e $\hat{\theta}_0^{A'}$ e $\hat{\theta}_0^{B'}$ subvectores das $(k\times 1)$ componentes identificáveis das subamostras A e B.

Em seguida através de uma combinação (convexa) numa média ponderada constrói-se o subvector:

$$\hat{\theta}_0^{AB} = \frac{1}{\sqrt{2}}\hat{\theta}_0^A + \left(1 - \frac{1}{\sqrt{2}}\right)\hat{\theta}_0^B.$$

Pondo de lado a subamostra A, retiram-se alguma(s) alternativa(s) à subamostra B de modo a definir o subconjunto $\tilde{C}_{N^B}$, e estima-se o subvector $\hat{\beta}_1^B = (\hat{\gamma}_1^B, \hat{\theta}_1^B)$ através da maximização do logaritmo da função de verosimilhança correspondente, ou seja:

$$Máx.\ \ln L_1^B(\beta) = \sum_{j \in N^B} Y_i \ln P(i|\tilde{C}_{N^B}).$$

Sob a hipótese de que a probabilidade de escolha individual é logit multivariada e que as características individuais são amostras aleatórias de uma dada distibuição subjacente, pode construir-se a razão de verosimilhanças que compara o estimador com restrições (nomeadamente a IAI) com o estimador da máxima verosimilhança, que corresponde ao teste:

$$SH = -2\left[\ln L_1^B(\hat{\theta}_0^{AB}) - \ln L_1^B(\hat{\theta}_1^B)\right]$$

seguindo uma distribuição assimptótica χ^2, com k graus de liberdade (i.e., a dimensão dos subvectores $\hat{\theta}_0^A$, $\hat{\theta}_0^B$, $\hat{\theta}_1^B$, $\hat{\theta}_0^{AB}$).

Alguns problemas surgem com estes tipos de testes apresentados :

Primeiro, quanto ao teste HM, não é obrigatório ser positivo para amostras finitas.

Segundo, nem todos os elementos de β são identificados pelo subconjunto $\tilde{C}$, o que implica que a estatística dos testes tem de ser calculada pela componente identificável de β.

E por último, não é claro como se escolhe as alternativas de C que formam o subconjunto $\tilde{C}$.

4.3. Testes Indirectos à IAI

4.3.1. Introdução

Tal como foi referido, nesta secção considera-se um modelo alternativo específico no qual se baseia o teste.

Dada uma hipótese alternativa paramétrica, podemos apelar aos testes convencionais para restrições de parâmetros: ensaio de Wald, teste da razão de verosimilhanças e do multiplicador de Lagrange. Pois é sabido, que para desvios locais à hipótese nula, estes testes têm boas propriedades para grandes amostras [ver Cox e Hinckley (1974)].

Já que o modelo logit é caracterizado pela propriedade IAI sse V_i só depender de X_i, a não se verificar esta restrição qualquer modelo de escolha discreta pode ser expresso na forma de um logit, o que nos leva a concluir que o teste mais directo e simples a realizar é contra o logit universal, ou seja, testar se as varíaveis de X_j , $j \neq i$ foram omitidas erradamente da especificação de V_i .

O modelo logit universal mantém a forma do logit multinomial ao pôr cada razão de probabilidades como uma função das características de todas as alternativas, o que implica que não goze da propriedade da independência das alternativas irrelevantes.

Pode ser escrito como:

$$P(i) = \frac{\exp T_i\beta}{\sum_j \exp T_j\beta}$$

em que os T_i são função de todas as variáveis explicativas do modelo, isto é, de todos os X_j e não apenas dos atributos da alternativa i, podendo tomar uma forma não linear, o que implica uma estimação mais complexa.

Por outro lado, é difícil dar uma interpretação económica (apenas pode ser descrito como um padrão de elasticidades cruzadas), já que não é consistente com a maximização da utilidade, sendo apenas uma aproximação flexível a uma forma funcional geral.

O modelo logit multinomial é um caso particular, que se verifica, quando a forma da componente sistemática da função utilidade só depende das características da alternativa a que respeita.

Se bem que este teste possa ser ensaiado com qualquer um dos três procedimentos atrás referidos, é de referir que o teste de Wald e o da razão de verosimilhanças exigem a estimação do logit universal, o que pode ser insuperável se tivermos um número demasiado elevado de variáveis explicativas.

Logo opta-se pelo teste do multiplicador de Lagrange [ver McFadden (1987)]. Opção tomada igualmente nos dois testes seguintes, já que não necessita de estimação dos parâmetros do modelo alternativo.

Contudo há que ter cuidado porque a rejeição do MLM não implica necessariamente a validade da especificação do logit universal, já que a rejeição da hipótese nula pode ser devido a outra má especificação que não a omissão de variáveis.

4.3.2. Hipótese Alternativa: Modelo Logit Sequencial

Consiste numa sobreposição de dois modelos logit a diferentes niveís, isto é, tem uma estrutura de árvore (Modelo de Escolha Hierárquica, Sequencial ou Encadeada)[10]. As alternativas são semelhantes dentro de cada nível e diferentes de nível para nível. Cada nível é especificado segundo um modelo logit.

Enquanto no modelo logit multinomial os erros $(\varepsilon_1,\ldots,\varepsilon_J)$ têm distribuições de Valor Extremo Independente, no modelo logit sequencial a distribuição conjunta dos resíduos segue uma distribuição de Valor Extremo Generalizado [ver Maddala (1983) e Chow (1983)]:

$$F(\varepsilon_1,\ldots,\varepsilon_J)=\exp\left\{-\sum_{i=1}^{q}\left[\sum_{j\in C_i}\exp\left(-\frac{\varepsilon_j}{\rho_i}\right)\right]^{\rho_i}\right\} \tag{10}$$

onde q representa os subconjuntos do conjunto de escolha $C_1,\ldots,C_q$, divididos de acordo com o grau de substituição entre as alternativas.

O ρ representa uma medida de associação entre as variáveis, sendo o coeficiente de correlação definido por $(1-\rho^2)$, com $\rho=1$ dentro de cada nível e $0<\rho<1$ entre os níveis.

---

[10]Por exemplo num modelo com três níveis, as J alternativas no nível 1 podem ser agrupadas nas J_2 alternativas do nível 2 e por sua vez agrupadas nas J_3 alternativas do grupo 3.

No caso de $q = 1$ e $\rho_i = 1 \quad \forall i$, o modelo logit sequencial fica reduzido ao logit multinomial .

O que possibilita um padrão geral de dependência entre as escolhas, evitando assim o problema ilustrado pelo Paradoxo autocarro azul/ autocarro vermelho (propriedade IAI), que era no fundo a principal restrição do modelo logit.

Demonstra-se que, se os resíduos forem distribuídos de acordo com (10) então a probabilidade da alternativa i ser escolhida pode ser definida como:

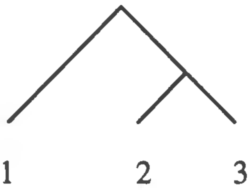
$$P(i) = \frac{\exp V_i \cdot G_i}{G(.)} = \frac{\exp\{V_i + \ln[G_i(.)]\}}{\sum_{j=1}^{J} \exp\{V_j + \ln[G_j(.)]\}} \qquad (11)$$

onde

$$G(.) = \sum_{i=1}^{M} \left[\sum_{j \in C_i} \exp\left(-\frac{\varepsilon_j}{\rho_i} \right) \right]^{\rho_i}$$

em que G_i é a derivada de G em ordem a $\exp V_i$.

Considerando o caso de três alternativas, o qual vai ser objecto de estudo na parte prática desta tese, e um q=2, pode-se, por exemplo, agrupar as alternativas do seguinte modo[11]:



$$F(\varepsilon_1,\varepsilon_2,\varepsilon_3)=\exp\{-G[\exp(-\varepsilon_1),\exp(-\varepsilon_2),\exp(-\varepsilon_3)]\}$$

$$G(.)=\left\{\exp(\varepsilon_1)+\left[\exp\left(-\frac{\varepsilon_2}{\rho_1}\right)+\exp\left(-\frac{\varepsilon_3}{\rho 1}\right)\right]^{\rho_1}\right\}$$

$$P(i)=\frac{\exp V_i\cdot G_i(\exp V_1,\exp V_2,\exp V_3)}{G(\exp V_1,\exp V_2,\exp V_3)}$$

Como $G_i=\dfrac{\partial\, G(.)}{\partial \exp V_i}$, então:

$$G_1=0$$

$$G_2=\exp\left[V_2\left(\frac{1}{\rho-1}\right)\right]\left[\exp\left(\frac{V_2}{\rho}\right)+\exp\left(\frac{V_3}{\rho}\right)\right]^{(\rho-1)}$$

$$G_3=\exp\left[V_3\left(\frac{1}{\rho-1}\right)\right]\left[\exp\left(\frac{V_2}{\rho}\right)+\exp\left(\frac{V_3}{\rho}\right)\right]^{(\rho-1)}$$

---

[11] No estudo de simulação realizado foram consideradas, para além desta, mais duas possibilidades: [(1,3) e 2] e [(1,2) e 3].

Concretizando a expressão (11) vem:

$$P(1) = \frac{\exp V_1}{\exp V_1 + \exp\left(\frac{V_2}{\rho} + Q\right) + \exp\left(\frac{V_3}{\rho} + Q\right)}$$

$$P(2) = \frac{\exp\left(\frac{V_2}{\rho} + Q\right)}{\exp V_1 + \exp\left(\frac{V_2}{\rho} + Q\right) + \exp\left(\frac{V_3}{\rho} + Q\right)}$$

$$P(3) = \frac{\exp\left(\frac{V_3}{\rho} + Q\right)}{\exp V_1 + \exp\left(\frac{V_2}{\rho} + Q\right) + \exp\left(\frac{V_3}{\rho} + Q\right)}$$

em que $Q = (\rho - 1)\log\left[\exp\left(\frac{V_2}{\rho}\right) + \exp\left(\frac{V_3}{\rho}\right)\right]$

A sua estimação pode ser feita através do método da máxima verosimilhança [ver Amemiya (1985)] ou de um método de estimação sequencial [ver Borsch-Supan (1987)].

O modelo logit sequencial é útil quando um conjunto de utilidades pode ser naturalmente classificado em gupos independentes, permitindo correlações não nulas entre as utilidades de cada nível.

É aplicado a modelos com um número elevado de variáveis explicativas, em que o processo de decisão tem uma estrutura de árvore (pode ser alargado a casos de vários ramos da árvore).

Aparecendo como um "ajuste" entre a flexibilidade funcional (handicap do logit multinomial devido à propriedade IAI) e a facilidade de computação (fraqueza do modelo probit multinomial quando há um número elevado de alternativas). Contudo continua a pressupôr a homocedasticidade entre os erros, tal como o modelo logit.

No fundo, o modelo logit sequencial é uma generalização do modelo logit multinomial com a adição de algumas variáveis extra, incluídas na componente sistemática da função utilidade como se depreende da expressão (11).

Assim uma hipótese possível, é testar se os coeficientes adicionais são iguais a zero. O que pode ser feito através do teste do multiplicador de Lagrange para variáveis omitidas [ver McFadden (1987)].

Contudo, a abordagem usual é testar directamente se os parâmetros ρ_i são todos iguais a um. Sob a hipótese nula, o teste segue uma distribuição χ^2 com M graus de liberdade (número de coeficientes que se vai testar serem iguais a zero) [(ver Hausman e McFadden (1984)].

Se bem que McFadden (1987) mostre que este tipo de teste possa ser encarado como um teste de omissão de determinadas variáveis. No fundo testar $\rho = 1$ é testar a exclusão de

$$\left\{\log\left[\exp\left(\frac{V_2}{\rho}\right)+\exp\left(\frac{V_3}{\rho}\right)\right]-\frac{V_i}{\rho}\right\} = -\log\left[\frac{\exp\left(\frac{V_i}{\rho}\right)}{\exp\left(\frac{V_2}{\rho}\right)+\exp\left(\frac{V_3}{\rho}\right)}\right] \text{ de } V_i \quad \text{com } i = 1,2.$$

O teste do multiplicador de Lagrange é calculado com base numa regressão auxiliar, na qual se pretende testar a significância de alguns regressores.

A estatística do teste é, então, definida como $S = (N - T)R^2$, em que o é o coeficiente de determinação múltipla da regrassão auxiliar. Esta estatística tem distribuição χ^2 com M graus de liberdade, onde N é o número de observações da regessão auxiliar. Neste caso, com três alternativas, vem $S = 2NR^2$ seguindo uma distribuição χ^2 com um grau de liberdade.

Testar a exclusão de V_i dos $P(i)$ é equivalente a testar a exclusão dos W_i da regressão auxiliar.

4.3. Hipótese Alternativa: Modelo Dogit

O modelo dogit, também denominado por modelo logit modificado [ver Hensher e Johnson (1988)], é suficientemente flexível para evitar a imposição da restrição da propriedade IAI para todos os pares de alternativas, sem contudo perder o aspecto prático e intuitivo do formato do modelo logit.

Ou seja, este modelo evita ou "foge" ao dilema de que, à priori, o decisor seja obrigado a escolher entre um modelo submetido à restrição da IAI e outro que a exclua.

Proposto por Gaudry e Dagenais (1978), assenta numa formulação alternativa do logit, com hipóteses menos restritivas.

Pode ser especificado como:

$$P(i) = \frac{\exp V_i + \theta_i \sum_{j=2}^{J} \exp V_j}{\left(1 + \sum_j \theta_j\right) \sum_{j=1}^{J} \exp V_j} \tag{13}$$

em que $P(i)$ [12] é a probabilidade de escolher a i-ésima alternativa e θ_i um parâmetro não negativo associado à i-ésima alternativa.

Como já foi referido, um dos aspectos mais importantes e úteis deste modelo é o seu " escape" à propriedade IAI (pelo menos para alguns pares de alternativas), que pode ser demonstrado do seguinte modo:

No modelo apresentado em (12), as probabilidades relativas de duas quaisquer alternativas i e m podem expressar-se como:

$$\frac{P(i|C)}{P(m|C)} = \frac{\exp V_i + \theta_i \sum_{j=1}^{J} \exp V_j}{\exp V_m + \theta_m \sum_{j=1}^{J} \exp V_j}$$

onde obviamente esta razão depende das características de todas as alternativas.

Implica que, se os θ_i e θ_m forem diferentes de zero para todas as alternativas, a introdução de uma nova alternativa alterará a razão das probabilidades das mesmas.

Para que tal não aconteça, apesar de uma alteração no valor de uma característica da l-ésima alternativa (para $l \neq i,m$) ou da introdução (ou exclusão) de uma nova alternativa, é necessário que:

12 Esta expressão está definida de tal forma que verifica as condições para uma distribuição em probabilidade:

$$0 \leq P(i) \leq 1 \qquad i=1,\dots,N \qquad \sum_j P(j)=1$$

$$\frac{\exp V_i + \theta_i \sum_{j=1}^{J} \exp V_j}{\exp V_m + \theta_m \sum_{j=1}^{J} \exp V_j} = \frac{\exp V_i + \theta_i \sum_{j=1}^{J} \exp V_j^*}{\exp V_m + \theta_m \sum_{j=1}^{J} \exp V_j^*}$$

onde V_j^* representa o novo conjunto de escolha que sofreu alguma(s) das alterações acima referidas.

E tal só se verifica perante uma das três situações:

(a) $\theta_i = \theta_m = 0$

(b) $\frac{\theta_i}{\theta_m} = \frac{\exp V_i}{\exp V_m}$

(c) $\sum_j \exp V_j = \sum_j \exp V_j^*$

Se se verificar a condição (a) para todos os pares de alternativas, a expressão (12) reduz-se ao modelo logit multinomial.

Assim, esta formulação do modelo dogit permite que subconjuntos de probabilidades relativas sejam determinadas consistentemente com o axioma da IAI, enquanto que outros não. Logo o modelo é considerado flexível, no sentido em que não requer que todas as probabilidades relativas sejam determinadas da mesma maneira.

Ao mesmo tempo que Gaudry e Dagenais desenvolviam este modelo, Ben-Akiva e Swait (1985) examinavam o problema do processo de geração do conjunto de escolha, e um dos casos analisados levou a uma interpretação do modelo dogit em termos de "ser cativo" relativamente a uma alternativa:

Supõe-se que um indivíduo ou é cativo à alternativa $i \in C$ (conjunto total de escolha disponível - geralmente diferenciado entre os indivíduos) ou é livre de escolher entre todas as alternativas do conjunto C de acordo com um modelo logit multinomial.

Então a probabilidade de um indivíduo escolher a alternativa i, pode ser expressa como:

$$P(i|C) = \frac{u_i}{1+u} \times 1 + \frac{1}{1+u} \times \frac{\exp V_i}{\sum_j \exp V_j} \qquad (13)$$

em que:

$\frac{u_i}{1+\sum u_j} = \frac{u_i}{1+u}$ representa a probabilidade do indivíduo ser cativo à alternativa i e $\frac{1}{1+u}$ a probabilidade do decisor ser livre para escolher uma alternativa do conjunto C.

A expressão (13) pode ser escrita como:

$$P(i|C) = \frac{\exp V_i + u_i \sum_{j=1}^{J} \exp V_j}{(1+u)\sum_{j=1}^{J} \exp V_j}$$

Expressão esta, que é análoga à (12) proposta por Gaudry e Dagenais fazendo $\theta_i = u_i$.

Logo o θ_i no modelo dogit pode ser interpretado como o grau de "captação" na decisão de escolha, na medida que quanto mais pequeno for o θ_i menor é a probabilidade o indivíduo ser cativo à alternativa i.

Apesar do modelo logit multinomial estar incluído em (13), i.e., Gaudry e Dagenais mostram que, o modelo dogit não "obriga" à propriedade IAI inerente a esse tipo de modelos.

Na verdade este "escape" parece ter sido a principal motivação para desenvolver-se tal especificação, mais do que a interpretação de geração do conjunto de escolha adoptado por Ben-Akiva (1977).

Tal como foi dito atrás, o modelo dogit é pois um caso particular dos modelos de Conjunto de Escolha Variável desenvolvidos por Ben-Akiva e Swait (The Parametrized Logit Captivity Model- PCL).

Este tipo de modelos o que faz, é parametrizar os parâmetros u_i como funções de variáveis independentes, i.e., desde que u_i sejam não negativos, Ben-Akiva sugeriu que:

$$u(X_i) = \exp(DX_i) \qquad \forall i \in C \tag{14}$$

onde $D = (d_1, d_2, \dots, d_k)$ é um vector de parâmetros e X_i é um vector de características sócio-económicas do decisor e características da alternativa i.

Substituindo a expressão (14) na (12) resulta na especificação dos Modelos de Escolha Variável (PLC):

$$P(i|C) = \frac{\exp(DX_i)}{1+\sum_{j\in C}\exp(DX_j)} + \frac{1}{1+\sum_{j\in C}\exp(DX_j)} \times \frac{\exp(BY_i)}{\sum_{j\in C}\exp(BY_j)} \qquad \forall \mathrm{i} \in \mathrm{C}$$

$B = (b_1, \ldots, b_l, \ldots, b_L)$ é um vector de parâmetros do MLM e Y_i é um vector de características sócio-económicas do indivíduo e características da alternativa i.

Gaudry e Dagenais (1979) deram uma interpretação alternativa de θ_i como indíce do grau de influência das características de todas as alternativas na escolha da i-ésima alternativa.

Funcionando como um tipo de efeito rendimento que vem reforçar ou enfraquecer o efeito de substituição automático inerente aos MLM, na equação (12), através do cálculo das elasticidades directas para o modelo dogit:

$$E^{P(i)}_{X_i} = \frac{\partial P(i)}{\partial X_i} \cdot \frac{X_i}{P(i)} = \beta X_i \left[\frac{\exp V_i (1+\theta_i)}{\exp V_i + \theta_i \sum_j \exp V_j} - \frac{\exp V_i}{\sum_j \exp V_j} \right]$$

que representa a variação percentual na probalilidade, induzida por uma variação de um por cento no valor de uma das características da i-ésima alternativa X_i (se é linear nos parâmetrosβ e nas variáveis X_i).

Se $\theta_i = 0$ então esta expressão reduz-se à já conhecida elasticidade do modelo logit multinomial:

$$E^{P(i)}_{X_{ik}} = \beta X_{ik} \left[1 - \frac{\exp V_i}{\sum_j \exp V_j} \right]$$

Para isolar o efeito do θ_i divide-se a expressão por $\exp V_i$, vindo então:

$$E_{X_{ik}}^{P(i)} = \beta_k X_{ik} \left[\frac{1+\theta_i}{1+\theta_i \left(\dfrac{\exp V_i}{\sum_j \exp V_j} \right)} - \frac{\exp V_i}{\sum_j \exp V_j} \right]$$

O primeiro termo dentro de parênteses é claramente menor do que 1, desde que $\sum_j \exp V_j \big/ \exp V_i$ seja sempre positivo e maior do que 1.

Assim a elasticidade será sempre menor que no caso do logit, então o θ_i é uma medida do efeito rendimento que reduz o efeito substituição, sem contudo anulá-lo, alterando assim o sinal da elasticidade.

Do mesmo modo, analisando a elasticidade cruzada :

$$E_{X_{ik}}^{P(i)} = \beta X_{ik} \left[\frac{\theta_i}{\dfrac{\exp V_i}{\exp V_j} + \theta_i \left(\dfrac{\sum_j \exp V_j}{\exp V_i} \right)} - \frac{\exp V_i}{\sum_j \exp V_j} \right]$$

Como se pode observar, a primeira parcela dentro de parênteses é positiva (no caso do MLM é nula, já que $\theta_i = 0$). Logo o efeito do dogit é aumentar o efeito cruzado comparado com a especificação do modelo logit. Contudo, pode ser demonstrado que a soma entre parêntesis é sempre negativa, e então as alternativas deveriam, tal como no logit, actuar como substitutos.

Como o MLM é um caso particular do modelo dogit, quando alguns parâmetros são iguais a zero, Tse (1987) derivou o teste multiplicador de Lagrange (ML) sob a hipótese nula do modelo logit (i.e., verificar-se a IAI) e usando o modelo dogit como hipótese alternativa.

A forma quadrática deste teste requer a componente do vector lagrangiano em relação aos parâmetros $\hat{\theta}$ e a matriz das covariâncias assimptótica inversa:

$$ML = s(\theta)'V(\theta)^{-1}s(\theta)$$

Sob a hipótese nula o teste segue uma distribuição χ^2 com J graus de liberdade (ou seja, o número de parâmetros θ_j).

As expressões para $s(\theta)$ e $V(\theta)$ podem ser encontradas em Tse (1987,p.284).

4.3.4. Hipótese Alternativa: Modelo Logit Heterogéneo

Assenta na hipótese de que o modelo tem variações adicionais não controladas em relação às utilidades, ou seja, a utilidade associada à alternativa i não depende só do vector X_i (vector das características das alternativas observadas) e de ε_i, mas também de outra variavél η_i.

Então, podemos expressar a utilidade da alternativa i como:

$$U_i = X_i\beta + \eta_i + \varepsilon_i$$

Em que η_i podem ser vistas como variáveis omitidas que deveriam estar incluídas no vector X_i se os dados estivessem disponíveis aquando a especificação do modelo.

Outra possível interpretação é supôr que são variáveis aleatórias que resultam de variações aleatórias dos parâmetros.

E assim consegue-se superar a restrição da IAI ao ter em conta a existência de heterogeneidade (em relação aos indivíduos) no modelo de utilidade aleatória, ou seja, as variáveis aleatórias podem ser correlacionadas de alternativa para alternativa.

Este modelo, tal como o logit multinomial, continua a ter por hipótese que os erros ε_i são variáveis aleatórias independentes e identicamente distribuidas seguindo uma distribuição Valor Extremo Tipo I, sendo independentes de X e η e é igualmente coerente com o princípio da maximização da utilidade.

Onde $\mathrm{cov}(\omega_i, \omega_j) = \omega_{ij}$ e a correspondente matriz das covariâncias é dada por $\Omega = [\omega_{ij}]$. Esta matriz determina a amplitude do efeito de η sobre as utilidades estocásticas, e por sua vez, sobre as probabilidades de escolha. Se $\Omega = 0$, aplica-se o modelo logit multinomial convencional, pois não há efeito na probabilidade de escolha. Pelo contrário, se $\Omega \neq 0$ então a probabilidade de um determinado indivíduo escolher a alternativa i é dada por:

$$\tilde{P}(i|X,\eta) = \frac{\exp(X_i\beta + \eta_i)}{\sum_{j=1}^{J} \exp(X_j\beta + \eta_j)}$$

A probabilidade de escolha que é aplicável para estimar β, quando só os dados de X estão disponíveis é:

$$P(i|X) = \int \frac{\exp(X_i\beta + \eta_i)}{\sum_{j=1}^{J} \exp(X_j\beta + \eta_j)} f(\eta|X)\, d\eta$$

em que $f(\eta\,|X)$ corresponde à função de densidade de η dado X.

Acontece que esta função é desconhecida, logo há que fazer algum pressuposto:

- Uma primeira alternativa é considerar que se conhece a forma da distribuição (abordagem paramétrica). Por exemplo, se por hipótese, $f(\eta\,|X)$ segue uma distribuição Gama, obtém-se o modelo Gama-Logistíco [ver Steckel e Vanhonacker (1988)]; outra hipótese, é a aproximação paramétrica de Tse (1989) que, baseando-se numa versão modificada da distribuição Gama [ver Johnson e Kotz (1972)], apresenta o modelo de utilidade proporcional.
- Uma segunda alternativa é considerar uma especificação semiparamétrica, usando uma distribuição discreta de mistura, conduzindo a um modelo de ponto de suporte com massa. Contudo quando o número de alternativas é elevado a sua estimação é bastante complexa.
- E por último, a alternativa não paramétrica (e a que vai ser objecto de estudo desta tese) aplicada por Chesher e Santos Silva (1992), em que $f(\eta\,|X)$ é desconhecida, ao desenvolverem uma aproximação quando a variância é pequena para as probabilidades de escolha de cada alternativa.

Para obter a aproximação para as probabilidades do modelo heterogéneo, desenvolve-se $\tilde{P}(i|X,\eta)$ em série de Taylor de segunda ordem em torno de η_i, integrando termo a termo e por fim normalizando através de funções de X e β, $r_i^{st}(X,\beta)$, de modo a que as probabilidades pertençam ao intervalo [0,1].

Pode-se então definir o modelo logit multinomial alargado, no qual a probabilidade de a alternativa i ser escolhida é expressa como:

$$g(i|X,\beta,\Omega)=\frac{\exp[X_i\beta+\omega_{st}r_i^{st}(X,\beta)]}{\sum_{j=1}^{J}[\exp(X_j\beta+\omega_{st}r_j^{st}(X,\beta)]} \tag{15}$$

em que:

$$r_i^{st}(X,\beta)=0.5\tilde{P}^{st}(i|X)\,/\,P(i|X,\beta)$$

O efeito de primeira ordem da variação da alternativa não controlada nas componentes de utilidade não observadas induz uma série de funções de características não lineares, não estocásticas nas funções de índice. Levando a caracteristícas de muitas alternativas para dentro da função indíce de uma alternativa. O que implica que o modelo já não goze da propriedade IAI.

Este modelo é um membro não linear da classe dos modelos logit universal, nos quais as caracteristícas de todas as alternativas podem afectar a função indíce associada a cada escolha.

Contudo, a principal vantagem é a sua facilidade de ser interpretado e estimado desde que os parâmetros tenham uma interpretação estrutural, isto é, os ω_{st} como variâncias e covariâncias das componentes de utilidade não observadas associadas às alternativas; os β como os parâmetros de um modelo de probabilidades de escolha condicional a um conjunto alargado de características de alternativas.

Dado que $r_i^{st}(X,\beta)$ é um termo não linear na função índice de $g(i|X,\beta,\Omega)$, a estimação pode tornar-se um pouco lenta. Contudo pode-se simplificar a aproximação a $\overline{P}(i|X)$ estimando r_i^{st} não em β, mas noutro vector de parâmetros, $b(\Omega)$, tal que $[\beta - b(\Omega)] \to 0$ à medida que $\Omega \to 0$.

O modelo de escolha discreta pode, então, ser escrito como:

$$d(i|X,\beta,b,\Omega) = \frac{\exp[X_i\beta + \omega_{st} r_i^{st}(X,b)]}{\sum_{j=1}^{J}[\exp(X_j\beta + \omega_{st} r_j^{st}(X,b)]} \qquad (16)$$

sendo $r_i^{st}(X,b)$ as novas variáveis adicionadas à especificação do logit original, de modo a considerar o efeito de primeira ordem da heterogeneidade não observada na função utilidade.

A vantagem desta expressão sobre a anterior, é que os termos não lineares são eliminados das funções indíce e por isso a estimação do modelo é bastante simplificada.

Contudo, a matriz das covariâncias derivada da matriz de informação com base na expressão (16), sendo b fixo, é incorrecta, visto que ignora a variabilidade das estimativas de b. Contudo esta matriz das covariâncias é correcta até à ordem da aproximação considerada.

Excepto em determinados casos, na prática, depois de estimar $\tilde{p}(i|X)$ através do modelo logit, constroem-se as funções $r_i^{st}(X,\beta)$ calculando β a partir das estimativas do logit, estima-se então a expressão (16) de modo a obter estimativas das probabilidades $\overline{p}(i|X)$ e os parâmetros β e Ω.

A base para construir testes para detectar desvios à especificação do logit é $\Omega = 0$, já que, quando tal se verifica, a expressão reduz-se ao modelo logit multinomial convencional.

Então a hipótese nula, $\Omega = 0$, é testada através dum teste do multiplicador de Lagrange desenvolvido por Chesher e Santos Silva (1992), baseado na diferença entre as covariâncias amostrais dos resíduos da forma $Y_i - p(i|X,\hat{\beta})$ e os valores esperados calculados sob a hipótese nula.

Este teste pode também ser considerado como pertencente à classe dos testes estatísticos da matriz de informação [ver White (1982)]: na medida em que as expressões (15) e (16) são aproximações ao modelo logit com termos independentes especifícos a cada alternativa, então $\Omega = 0$ é o teste matriz de informação que aparece quando os elementos desta matriz associados a estes parâmetros são considerados. É pois esta a interpretação do teste da matriz de informação como um teste para detectar variação do parâmetro aleatório [ver Chesher (1984)].

Um dos aspectos interessantes destes resultados é que o modelo logit alargado fornece um modo para calcular versões da razão de verosimilhanças do teste estatístico da matriz de informação.

5. ESTUDO DE MONTE CARLO

O objectivo deste estudo de simulação é pegar no melhor teste do artigo de Fry e Harris (1993) e compará-lo com outros dois que não foram objecto de análise, nomeadamente, o teste score contra o modelo logit sequencial e o teste da matriz de informação contra o modelo logit heterogéneo.

Para que se possam comparar resultados há que utilizar um processo de simulação idêntico, i.e., todas as hipóteses e a própria estrutura do modelo terão que ser as mesmas das do artigo em causa.

5.1. Descrição do estudo

Num estudo de Monte Carlo, o investigador especifica um modelo estatístico teórico subjacente a um processo amostral, gera dados da amostra consistentes com este processo (números aleatórios), desenvolve estimativas dos parâmetros desconhecidos consistentes com uma ou mais regras, e analisa as estimativas de modo a determinar as características amostrais.

Neste tipo de estudos os resultados obtidos dependem, então, da experiência considerada. Como tal, inicialmente há que descrever todo o processo de simulação que foi utilizado para estudar os três testes escolhidos: teste de Small-Hsiao, teste do multiplicador de Lagrange (hipótese alternativa: modelo logit sequencial) e teste da matriz de informação - na versão razão de verosimilhanças - (hipótese alternativa: modelo logit heterógeneo).

Em todas as experiências efectuadas o número de alternativas considerado foram três, isto é, J=3, a variante do modelo logit analisado foi o modelo logit multinomial em que o vector das variáveis explicativas X_i é constituido por uma constante e duas variáveis independentes, geradas através de uma distribuição normal standardizada. Uma vez gerado o vector X_i, permanece constante para todas as experiências de simulação, o que quer dizer que estamos em presença do modelo logit multinomial:

$$P(i|C) = \frac{\exp X\beta_i}{\sum_{j=1}^{J} \exp X\beta_j}$$

Em todas as experiências considerou-se três tamanhos amostrais: 250, 500, 1000. Para os cálculos utilizou-se o package TSP (Times Series Processor), versão 4.2.

O primeiro conjunto de simulações foi esboçado com o objectivo de investigar as propriedades dimensionais dos testes: o processo de geração subjacente para os dados observados foi assumir um modelo de maximização de utilidade aleatória, em que os erros, ε_j's, foram gerados sob a hipótese de serem independentes, identicamente distríbuidos seguindo uma distribuição Gumbel.

A escolha dos vectores β_j (j=1,2,3), tendo presente a simplicidade em termos de resultados, foi a seguinte: $\beta_1 = (\gamma \ \ -1 \ \ 1)$, $\beta_2 = (\gamma \ \ 1 \ \ -1)$ e $\beta_3 = (0 \ \ 0 \ \ 0)'$, onde γ é a constante de Euler, $\gamma = 0.577216$. Note-se que β_3 está normalizado a zero para se poder identificar o modelo, tal como já foi discutido atrás no ponto 3.

Foram consideradas 5000 e 2000 réplicas para analisar a dimensão e potência dos testes, respectivamente[13]. Para cada réplica, os ε_j's são adicionados aos conhecidos V_j's obtendo-se os U_j's. O índice da variável observável Y_i é igual a 1,2 ou 3 consoante a alternativa que maximiza a utilidade, U_j (j=1,2,3).

O modelo logit multinomial é então estimado, e os três testes são realizados com um nível de significância de 5%.

Os resultados obtidos encontram-se no quadro 1. Convém notar que no teste de Small-Hsiao consideraram-se os três subconjuntos possíveis: SHR1 (retirou-se a alternativa 1) e do mesmo modo, SHR2 e o SHR3.

No caso do teste do multiplicador de Lagrange contra o modelo logit sequencial desconhecendo-se quais as alternativas semelhantes dentro do nível, colocaram-se as três hipóteses possíveis: LMR1 (as alternativas 2 e 3 são semelhantes), LMR2 (as alternativas 1 e 3 são semelhantes) e LMR3 (as alternativas 1 e 2 são semelhantes). O LMR, por sua vez, testa simultâneamente as três hipóteses colocadas anteriormente.

Por fim, no teste da matriz de informação contra o modelo logit heterógeneo, o IMR1 e o IMR2 correspondem às variâncias do termo independente da primeira e da segunda alternativas e o IMR3 à covariância entre elas. O IMR é o teste conjunto,ou seja é um teste à matriz $[\Omega] = \omega_{ij}$.

---

[13]Quanto mais precisa é a estimativa que se quer obter maior é o número de réplicas que se deve considerar.

No segundo conjunto de simulações pretende-se averiguar a potência dos testes, é pois necessário gerar dados Y_i's sob a hipótese alternativa, ou seja, de que não se verifique a propriedade IAI. A especificação considerada foi a do modelo probit multinomial (MPM). Os vectores β_j's mantêm os mesmos valores aquando a análise da dimensão dos testes. Neste modelo os ε_j's seguem uma distribuição normal multivariada com média 0 e matriz das covariâncias Σ_ε, esta matriz é definida da mesma maneira que Fry e Harris definiram no seu artigo de modo a poder fazer algumas comparações:

$$\Sigma_\varepsilon = \begin{bmatrix} 1 & \rho & 0 \\ \rho & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

onde ρ é o coeficiente de correlação entre as variáveis aleatórias das duas primeiras alternativas. Esta matriz é consistente com o modelo probit multinomial usado para calcular as probabilidades de rejeição em Tse (1987).

5.2. Análise dos resultados

Analisando a dimensão dos testes pelo quadro 1, podemos concluir que o teste de Small-Hsiao é o mais fraco. Se bem que pertença a uma classe de testes que têm o inconveniente de serem muito sensíveis relativamente à escolha das alternativas eliminadas do conjunto C, ou seja, contêm alguma arbitrariedade na selecção feita pelo investigador: uma ou mais alternativas têm de ser eliminadas do conjunto de escolha e o modelo logit multinomial deve voltar a ser estimado para o conjunto de escolha reduzido. Contudo é o teste que melhor se comporta comparado com o de Hausman e McFadden (HM) e o de McFadden, Tye e Train (MTT), estudados no artigo de Fry e Harris (1993).

Já o teste do modelo logit sequencial como alternativa ao logit multinomial (visto como um teste de omissão de variáveis) é o que apresenta melhores resultados, na medida em que todos os valores são relevantes.

Igualmente bem comportado é o teste que tem como hipótese alternativa ao modelo logit multinomial o logit heterógeneo ao apresentar apenas três valores não significativos.

QUADRO 1 - DIMENSÃO EMPÍRICA DOS TESTES PARA UMA DIMENSÃO NOMINAL DE 5 %

TESTES	$\alpha = 0.05$		
	n=250	n=500	n=1000
SHR1	0.075	0.059	0.059
SHR2	0.079	0.058	0.056 *
SHR3	0.077	0.063	0.059
LMR	0.046 *	0.047 *	0.049 *
LMR1	0.052 *	0.044 *	0.052 *
LMR2	0.047 *	0.047 *	0.053 *
LMR3	0.050 *	0.046 *	0.052 *
IMR	0.057 *	0.054 *	0.052 *
IMR1	0.043	0.043	0.049 *
IMR2	0.043	0.049 *	0.049 *
IMR3	0.052 *	0.040	0.052 *

* indica que a dimensão estimada é igual à dimensão nominal, ao nível de significância de 5%, calculada do seguinte modo:

$$0.05 \pm 1.96\sqrt{\frac{0.05 \times 0.95}{5000}}$$

Quanto à potência, definida como a probabilidade de rejeitar a hipótese nula dado que é falsa e tendo como valor óptimo 100 %, continua a ser o teste de Small-Hsiao o mais fraco, ou seja, o menos potente, tendo apenas valores relevantes em SHR3 para um correlação positiva, de 0.9.

Ao passo que os outros dois testes, para a mesma correlação, apresentam valores bastante próximos (ou iguais) a 100 %. Dado que, como foi visto atrás, no modelo logit sequencial o ρ varia entre 0 e 1, o teste não detecta correlações negativas, daí resulta valores tão baixos. Contudo, aumentando o n, o comportamento do teste melhora.

Já para correlações positivas o teste é bastante potente, comparado com o teste da matriz de informação, nomeadamente para $\rho = 0.9$.

Outro aspecto a retirar da análise do quadro 2, é que enquanto o teste do multiplicador de Lagrange é bastante potente quando se sabe quais as alternativas semelhantes (caso LMR1: 0.912, 0.998 e 1 para n=250, 500, 1000, respectivamente), quando se desconhece tal facto, é-se obrigado a fazer um teste conjunto, LMR, onde o seu comportamento piora um pouco.

Enquanto que, com o teste da matriz de informação passa-se precisamente o contrário, como teste conjunto (IMR) apresenta resultados mais satisfatórios do que quando testado separadamente (IMR1, IMR2, IMR3).

Por tudo isto, podemos considerar que ambos os testes são bastante potentes contra a alternativa do modelo probit multinomial.

QUADRO 2 - FREQUÊNCIAS DE REJEIÇÃO AO NÍVEL DE 5%
ALTERNATIVA MODELO PROBIT MULTINOMIAL

TESTES	$\rho=-0.9$			$\rho=-0.5$		
	n=250	**n=500**	**n=1000**	**n=250**	**n=500**	**n=1000**
SHR1	0.078	0.058	0.050	0.073	0.064	0.053
SHR2	0.078	0.056	0.044	0.073	0.057	0.050
SHR3	0.076	0.064	0.065	0.076	0.056	0.049
LMR	0.038	0.039	0.108	0.036	0.050	0.080
LMR1	0.059	0.086	0.166	0.052	0.067	0.100
LMR2	0.041	0.047	0.093	0.053	0.057	0.078
LMR3	0.035	0.055	0.079	0.041	0.044	0.064
IMR	0.086	0.106	0.182	0.079	0.089	0.138
IMR1	0.044	0.049	0.057	0.040	0.045	0.048
IMR2	0.046	0.049	0.056	0.039	0.051	0.049
IMR3	0.050	0.055	0.076	0.045	0.052	0.052

TESTES	$\rho=0$			$\rho=0.5$		
	n=250	**n=500**	**n=1000**	**n=250**	**n=500**	**n=1000**
SHR1	0.073	0.066	0.060	0.093	0.071	0.092
SHR2	0.071	0.059	0.062	0.093	0.079	0.072
SHR3	0.059	0.060	0.058	0.071	0.067	0.083
LMR	0.043	0.053	0.086	0.130	0.236	0.447
LMR1	0.060	0.054	0.054	0.190	0.339	0.547
LMR2	0.042	0.044	0.044	0.090	0.147	0.280
LMR3	0.041	0.043	0.040	0.081	0.159	0.252
IMR	0.079	0.104	0.125	0.131	0.206	0.377
IMR1	0.059	0.082	0.115	0.094	0.177	0.333
IMR2	0.053	0.083	0.120	0.108	0.162	0.329
IMR3	0.051	0.069	0.097	0.139	0.288	0.519

TESTES	$\rho=0.9$		
	n=250	**n=500**	**n=1000**
SHR1	0.127	0.102	0.096
SHR2	0.137	0.116	0.127
SHR3	0.304	0.567	0.911
LMR	0.762	0.980	0.999
LMR1	0.912	0.998	1.000
LMR2	0.589	0.925	0.996
LMR3	0.605	0.935	0.998
IMR	0.767	0.984	1.000
IMR1	0.206	0.401	0.694
IMR2	0.210	0.391	0.667
IMR3	0.676	0.977	1.000

6. CONCLUSÕES

Atendendo que o objectivo desta dissertação era fazer um estudo do comportamento de determinados testes de especificação em modelos logit multinomiais, tendo por base o estudo realizado por Fry e Harris (1993) sobre o mesmo assunto, procurando de algum modo explorá-lo e dar-lhe uma certa continuidade, começou por se expôr a teoria relativa a este tipo de modelos.

Assim sendo, apresentaram-se os modelos de escolha discreta, particularizando em seguida, o modelo logit multinomial, expondo-se as suas hipóteses e estrutura, bem como o método mais simples de estimação dos respectivos parâmetros: método da máxima verosimilhança.

Como não podia deixar de ser analisou-se a principal propriedade, e a que no fundo é testada, a propriedade da independência das alternativas irrelevantes, inerente somente a este tipo de modelos. Por último, dentro deste ponto, expôs-se o modelo probit multinomial, onde a diferença fundamental, em relação ao modelo logit multinomial, reside na hipótese que se faz acerca da distribuição dos erros.

Depois de feita esta abordagem, passou-se para o ponto principal deste estudo, ou seja, a exposição e desenvolvimento dos testes de especificação ilustrada com um estudo de simulação. Apesar de se terem apresentado vários testes, nem todos foram testados já que as conclusões a que chegaram Fry e Harris assim o justificaram.

Nomeadamente o teste de Hausman e McFadden, que apesar de ser dos mais popular, tem dimensão reduzida e é pouco potente, sendo bastante sensível à alternativa eliminada do conjunto C. Quanto aos outros testes baseados na partição do conjunto de escolha o único que apresenta resultados minimamente satisfatórios em termos de dimensão é o teste de Small-Hsiao, tendo sido por isso incluído no estudo de simulação realizado, sendo os resultados semelhantes aos do artigo acima referido.

Contudo, pode-se concluir que este tipo de testes são inúteis dado o seu comportamento tanto sob a hipótese nula como sob a alternativa.

Quanto ao teste tendo como hipótese alternativa o modelo dogit, não foi objecto de análise nesta tese, dado que pelo outro estudo, a conclusão a que se chega é que ele é apenas útil quando testado contra alternativas dogit. O que não é suficiente para avaliar a potência e dimensão dum teste.

Procurou-se assim, avançar com outros dois testes possíveis, não incluídos no artigo referido, o do multiplicador de Lagrange contra o modelo logit sequencial e o da matriz de informação contra o modelo logit heterogéneo. Estes, vieram-se a revelar testes bem comportados tanto em termos de dimensão como de potência.

Se tiver que se optar por um deles, tudo depende do tipo de informação disponível. Dispondo de informação acerca da correlação existente entre as alternativas, bem como as que são semelhantes, então opta-se pelo teste do multiplicador de Lagrange contra o modelo logit sequencial. Senão, o melhor é escolher o teste da matriz de informação contra o modelo logit heterogéneo.

Ou seja, o teste do multiplicador de Lagrange contra o modelo logit sequencial, revela-se muito potente em determinada direcção e pouco nas restantes. Ao passo que o teste da matriz de informação contra o modelo logit heterogéneo é relativamente potente em todas as direcções.

REFERÊNCIAS BIBLIOGRÁFICAS

Amemiya,T.(1981) - "Qualitative Response Models: A survey". *Journal of Economic Literature*, 19 (December): 1483-1536.

Amemiya,T. (1985) - *Advanced Econometrics*. Cambridge,MA,MIT Press.

Ben-Akiva, M. (1977) - *Choice Models with Simple Choice Set Generation Processes*. Working Paper. Department of Civil Engineering, MIT Cambridge, Mass.

Ben-Akiva, M.; Lerman, S.R. (1985) - *Discrete Choice Analysis*. Cambridge,MIT Press.

Borsch-Supan, A. (1987) - *Econometric Analysis of Discrete Choice*. Germany, SpringerVerlag.

Chesher, A.D. (1984) - " Testing for neglected heterogeneity". *Econometrica*, 52: 865-872.

Chesher, A.; Santos Silva, J. (1992) - *"Discrete Choice Models Without the IIA Property"*. Bristol Universty.

Chow, G.C. (1983) - *Econometrics*. Singapore,McGraw-Hill,Inc.

Clark, C. (1961) - " The Greatest of a Finite Set of Random Variables". *Operation Research* (March-April).

Cox, D.R.; Hincley, D. (1974) - *Theoretical Statistics*. London: Chapman and Hall.

Cramer, J.S.; Ridder, G. (1988) - "The Logit Model in Econometrics". *Statistica* and *Measuring the Determinants of Model Choice*.Henshere Q. Dalvi, eds. Teakfield, London.

Domencich, T.; Mc Fadden,D. (1975) - *Urban Travel Demand*. Amsterdam, North-Holland.

Fomby, T.B. et al. (1984) - *Advanced Econometrics Methods*. New-York, Springer-Verlag.

Fry, T.R.L.; Harris, M.N. (1993) - "*A Monte Carlo Study of Tests for the Independence of Irrelevant Alternatives Property*". Department of Econometrics,Monash University, Clayton, Vic. 3168, Australia

Gaudry, M.; Dagenais, M. (1979) - " The Dogit Model". *Transportation Research*, Part B - Methodological, 13, 105-111.

Hall, B.H.; Cummins C.; Schnake,R. (1991) - *Times Series Processor (TSP).* Version 4.2. Reference Manual

Hausman, J. (1978) - " Specification Tests in Econometric". *Econometrica*, 46: 1251-1271.

Hausman, J.; McFadden, D. (1984) - " Specification Tests for the Multinomial Logit Model". *Econometrica*, 52: 1219-1240.

Hendry, D.F. (1980) - "Econometrics: alchemy or science?". *Economica*, 47, 387-406.

Hensher D. A. ;Jonhson Lesler W. (1988) - *Applied Discrete- Choice Modelling*. Croom Helm London.

Horowitz, J. (1981a) - " Testing the Multinomial Logit Model against the Multinomial Probit Model eithout Estimating the Probit Prameters". *Transportation Science*, 15: 153-163.

Horowitz, J. (1981b) - " Identification and Diagnosis of Specification Errors in the Multinomial Logit Model". *Transportation Research*, 15B: 345-360.

Johnson N.L.; Kotz, S. (1970) - *Continuos Univariate Distributions- 1.* New York, John Wiley.

Johnson N.L.; Kotz, S. (1972) - *Distributions in Statistics: Continuos Multivariate Distributions.* Vol. 1 e 2 . New York : John Wiley

Lerman, S.; Manski, C. (1982) - " A Model of the Effect of Information Diffusion of Travel". *Transportation Science*, 16: 171-191.

Luce, R. (1959) - *Individual Choice of Behavior: A Theoretical Analysis*. Wiley, New York

Maddala, G. S. (1983) - *Limited Dependent and Qualitative Variables in Econometric*. New-York. Cambridge University Press.

Manski, C. (1973) - *"The Analysis of Qualitative Choice". Ph.D.dissertation*. Department of Economics, MIT, Cambridge. Mass

McFadden, D. (1974) - "Conditional Logit Analysis of Qualitative Choice Behavior". Zaembka, P. (ed). *Frontiers in Econometrics*. New York, Academic Press, pp 105-142.

McFadden,D. (1976) - "The Revealed Preferences of Public Bureaucracy". *Bell J 7*:55-72.

McFadden,D.; W. Tye; Train, K. (1976) - "An Aplication of Diagnostic Tests for the Indepedence from Irrelevant Alternatives Property of the Multinomial Logit Model". *Transportation Research Record*, 637: 39-46.

McFadden, D.(1981) - Econometric Models of Probabilistic Choice. Charles F. Manski and Daniel Mc Fadden, eds.,*Strutural analysis of discrete data with econometric aplications* (MIT Press,Cambridge,MA), 198-272.

McFadden, D. (1984) - "Qalitative Response Models". Griliches, Z.; Intriligator,M. (eds.). *Handbook of Econometrics*. vol.2. Amsterdam

McFadden, D. (1987) - " Regression-Based Specification Tests for the Multinomial Logit Model". *Journal of Econometrics*, 34: 63-82.

Silvey, D.S. (1959) - " The Lagrangian Multiplier Test". *Annales Math. Statist.* 30: 389-407.

Small, K.; Hsiao, C. (1985) - " Multinomial Logit Specification Tests". *International Economic Reviews*, 26: 619-627.

Steckel, J. H.; Vanhonacker, W. R. (1988) - " A Heterogeneous Conditional Logit Model of Choice. *Journal of Business & Economic Satistics*, 7: 61-65.

Swait,J.; Ben-Akiva, M. (1985) - "Empirical Test of a constrained choice discrete model: mode choice in S. Paulo. Brazil". *Transportation Research-B*. Vol. 21B: 103-115

Theil, H. (1969) - " A Multinomial Extention of thr Linear Logit Model". *International Economis Reviews*, 10: 251-259.

Tse, Y. K. (1987) - " A Diagnostic Test for the Multinomial Logit Model". *Journal of Business & Economic Statistics*, 7: 283-286.

Tse, Y. K. (1989) - " A Proportional Random Utility Approach to Qualitative Response Models". *Journal of Business & Economic Statistics*, 7: 61-65.

White, H. (1982) - " Maximum likelihood estimation of misspecified models". *Econometrica*, 50: 1-25.